Empirical Classical Force Fields
for Molecular Systems

P. H. Hünenberger & W. F. van Gunsteren

Laboratorium für Physikalische Chemie, ETH-Zentrum, CH-8092 Zürich

Contents

1 Introduction

With the continuing increase of the power of computers, the past decades have seen a rapid increase in the number, performance and accuracy of theoretical computational methods in chemistry (van Gunsteren et al., 1989 ff, Lipkowitz & Boyd, 1990 ff). One can distinguish three major classes of methods for the theoretical study of molecular properties, listed in order of decreasing computational expenses: (i) *ab initio* molecular-orbital methods (Hehre et al., 1986), (ii) semi-empirical molecular-orbital methods (Zerner, 1991), and (iii) empirical classical force-field methods. Since the available computing resources are most often the true limiting factor to numerical calculations, it has become clear that there is no universal method able to solve all possible problems, but that one should rather select the method that is the most suitable to a problem of interest. The properties of the observable(s) and system under consideration that will, together with the available computing power, largely determine which type of method can be used are (van Gunsteren & Berendsen, 1990): (i) the required system size, (ii) the required volume of conformational space that has to be searched or sampled (in terms of dynamics: the required time-scale), (iii) the required resolution in terms of particles (determined by the smallest entity, subatomic particle, atom, or group of atoms, treated explicitly in the model), and (iv) the required energetical accuracy of the interaction function. These requirements may be incompatible, in which case the observable cannot be computed adequately with the currently available computer resources (van Gunsteren et al., 1995b). When requirements (i) and (ii) are in conflict with requirement (iii), this conflict may be resolved by the design of hierarchical or hybrid models, where only the relevant degrees of freedom are treated with a more expensive, higher resolution method. This is often done, for example, in the study of acid- or base-catalysed, organic or enzymatic reactions in the bulk phase (Warshel, 1991, Field, 1993, Whitnell & Wilson, 1993, Liu et al., 1996a).

Molecular-orbital methods are well suited for the study of small molecules or small clusters of molecules (supermolecule) in vacuum (Keith & Frisch, 1994), or within an averaged solvent environment (Ángián, 1992, Cramer & Truhlar, 1992, 1994, Tomasi & Persico, 1994, Müller-Plathe & van Gunsteren, 1994), and give access to properties such as equilibrium geometries, vibrational frequencies, heats of formation, relative energies of conformers and isomerisation barriers. These problems are also addressed with an increasing accuracy by empirical methods (Bowen & Allinger, 1991, Dinur & Hagler, 1991, Maple et al., 1994a,b). Due to the size of the problem and volume of accessible conformational space, simulation of organic molecules or macromolecules in the condensed phase is the domain of atom-based empirical classical force fields (van Gunsteren & Berendsen, 1990). Long time-scale (or long relaxation time) problems involving large systems, such as protein folding or *de novo* protein design, can currently be addressed only by residue-based force fields (Gerber, 1992, Jones, 1994). Finding an accurate description of the interaction at this low particle resolution (*i.e.* a sufficient energetical resolution) is, however, a major difficulty.

Empirical classical force fields are based on a generalization of the Born-Oppenheimer approximation, that is, on an averaging of the quantum mechanical Hamiltonian over implicit degrees of freedom (electronic and possibly also of individual atoms) to obtain an analytical interaction function depending solely on the explicit degrees of freedom of the model. This averaging process occurs at three levels:

A. Averaging over the implicit degrees of freedom,

B. Averaging of a force-field term over the different chemical environments,

C. Averaging of a force-field term corresponding to an internal coordinate over the other force-field terms depending on the same coordinate.

Choosing the explicitly handled degrees of freedom and the interaction function are the two first steps in an empirical force field calculation. The third is the choice of a method to sample the conformational space (van Gunsteren et al., 1995a). This choice will also depend on the information required to compute the observable(s) of interest, namely:

A. Structural information:

Possible choices are systematic-search (for small size problems only), energy minimization (EM) to a local minimum, and enhanced search methods such as genetic algorithms, simulated annealing, use of a soft-core potential (Beutler et al., 1994), four-dimensional search (van Schaik et al., 1993, Beutler & van Gunsteren, 1994) or the local-elevation method (Huber et al., 1994) in molecular simulation.

B. Structural and thermodynamic information:

Methods of choice are Monte-Carlo (MC) sampling (Binder, 1992, Frenkel, 1993) or the dynamical simulation methods mentioned below.

C. Structural, thermodynamic and dynamical information:

In this case, equations of motion which explicitly contain time are required, such as the Schrödinger, Newton, Lagrange or Langevin equations of motion. Possible techniques are classical (van Gunsteren, 1993) or quantum (Field, 1993) molecular dynamics (MD) or stochastic dynamics (SD) simulations.

The present text will mainly concentrate on the choice of the explicit degrees of freedom to be included in the model, and the functional representation of the interaction function in atom and united-atom based force fields.

2 Choice of the explicit degrees of freedom

The choice of an elementary unit (*i. e.* the particle that will have no explicit internal degrees of freedom) is the first step in the design of an empirical classical force field. Possible alternatives for the elementary unit and explicitly treated degrees of freedom, together with the corresponding type of interaction function, are summarized in Table I. This choice will determine or strongly influence (van Gunsteren & Berendsen, 1990, van Gunsteren & Mark, 1992, van Gunsteren et al., 1995b):

A. The number of degrees of freedom that will have to be handled explicitly for describing a specific molecular system, and thus the computational effort.

B. The amount of conformational space that can be searched (or in terms of molecular dynamics, the reachable time-scale). Because available computing power is most often a limiting factor, for a system of a given size, the number of possible evaluations of the potential energy function will rapidly decrease with the number of explicit degrees of freedom.

C. The maximum resolution, in terms of particles (*e. g.* subatomic particles, atoms, group of atoms, or molecules) and processes (*e. g.* chemical reactions, conformational changes, ...) that can be achieved by the force field.

D. The type of functions that are likely to describe the interaction between elementary units in an adequate manner, that is, with a reasonable energetical accuracy.

E. The type of observables the force field may be able to describe correctly, and those which will necessarily stay inaccessible. Accessible observables will be those for which the extent of searchable conformational space (B), the force field resolution in terms of particles (C), and the force field accuracy (D) are sufficient.

Current developments in empirical classical force fields mainly follow five basic lines in terms of degrees of freedom (Bowen & Allinger, 1991, Dinur & Hagler, 1991, Gelin, 1993, Whitnell & Wilson, 1993, van Gunsteren et al., 1994, Jones, 1994), which will be described in subsections 2.1 – 2.5. Note that in 2.3 – 2.5, the number of explicit degrees of freedom is reduced essentially by decreasing the force field resolution in terms of particles. An alternative way to reduce the size of the conformational space to be searched is to limit the dimensionality or to discretize the coordinates (lattice methods, see *e. g.* Binder, 1992). These methods will not be discussed here.

2.1   Gas-phase force fields

The primary purpose of gas phase force fields is the accurate description of molecules in vacuum (Bowen & Allinger, 1991, Dinur & Hagler, 1991, Hagler & Ewig, 1994, Maple et al., 1988, 1994a,b, Hwang et al., 1994). These force fields may be used to either complete or replace more expensive *ab initio* molecular orbital calculations (Maple et al., 1994a), or to predict experimental gas-phase properties such as equilibrium geometries, vibrational frequencies, heats of formation, relative energies of conformers and energy barriers for isomerisation (Hwang et al., 1994). Rapid progress in the design of such force fields is made possible by (i) the absence of intermolecular forces, (ii) the increasing amount and reliability of data from *ab initio* molecular-orbital calculations and (iii) the use of systematic and relatively inexpensive procedures for parameter calibration using both theoretical and experimental data. These force fields, sometimes called class II force fields (Maple et al., 1994a,b), are usually characterized by a detailed description of covalent degrees of freedom, involving anharmonic (non-quadratic) potential energy terms and terms that couple the internal coordinates (non-diagonal energy terms). Typical examples are the force fields CFF (Lifson & Warshel, 1968, Warshel & Lifson, 1970, Lifson & Stern, 1981) and a recently modified version (Engelsen et al., 1995a,b), CVFF (Hagler et al., 1979a–c, Lifson et al., 1979), EFF93 (Dillen, 1995a,b), MM2 (Allinger, 1977, Bowen & Allinger, 1991), MM3 (Allinger et al., 1989, Lii & Allinger, 1989a,b, Bowen & Allinger, 1991) and QMFF/CFF93 (Maple et al., 1994a,b, Hwang et al., 1994).

The term gas-phase force field does not mean that such force fields cannot be extended for applications in condensed phase simulations. Experimental information on crystal structures is sometimes used in the parametrization procedure (Warshel & Lifson, 1970, Dillen, 1995b, Engelsen, 1995b). For real applications in liquid phase problems, however, these force fields will suffer from the same difficulties in parametrization as condensed-phase force fields (section 2.2), and whether the significantly improved accuracy gained in the gas phase by inclusion

of anharmonic and off-diagonal terms will result in a significant increase of accuracy in the simulated condensed-phase properties is still matter of discussion.

2.2 Condensed-phase force fields

The primary purpose of condensed-phase force fields is the accurate description of liquids, solutions of organic compounds or macromolecules (Allen & Tildesley, 1987, Brooks III et al., 1988, McCammon & Harvey, 1987, van Gunsteren & Berendsen, 1990) and crystals. Progress in the development of such force fields is slow, since (i) the dominant forces in the condensed phase are intermolecular forces which are not easily described and parametrized adequately, (ii) the relevance of data from *ab initio* molecular orbital calculations in vacuum (even when reaction-field corrections are applied) is limited, and the parametrization has to rely mostly on a small amount of experimental data concerning the condensed phase, and (iii) the design of systematic optimization procedures is in general not possible. One major reason for this impossibility is that the estimation of observables to be compared to experimental results generally requires a large number of evaluations of the potential energy function, and is therefore computationally expensive. In these force fields, the main effort is aimed at the description of non-bonded forces and torsional potential energy terms. Potential energy terms involving other covalent internal coordinates are often either quadratic-diagonal (so-called class I force field) or simply zeroed by the use of constraints. Typical examples are the force fields AMBER (Weiner & Kollman, 1981, Weiner et al., 1984, 1986, Pearlman et al., 1995), CHARMM (Brooks et al., 1983, Smith & Karplus, 1992, Nilsson & Karplus, 1986, MacKerell Jr. et al., 1995), CHARMm/QUANTA (Momany & Rone, 1992), DREIDING (Mayo et al., 1990), ECEPP/3 (Némethy et al., 1992), ENCAD (Levitt, 1983a,b, Levitt et al., 1995), EREF (Levitt, 1974), GROMOS (van Gunsteren & Berendsen, 1987), MAB (Gerber & Müller, 1995), OPLS(Jorgensen & Tirado-Rives, 1988), Tripos (Clark et al., 1989), UFF (Rappé et al., 1992) and YETI (Vedani, 1988).

2.3 Mean-solvent force fields

The purpose of a mean-solvent force field is the description of molecules in solution, but without an explicit treatment of the solvent degrees of freedom (van Gunsteren et al., 1994). Although an accurate description of the structure, mobility, dynamics and energetics of molecules in solution generally requires an explicit treatment of the solvent, the omission of all or almost all solvent degrees of freedom dramatically reduces the computational expenses, *e. g.* by a factor $10-50$ for biomolecules in solution. The explicit influence of the solvent is approximated here by its mean effect, and possibly also the effect of its mean fluctuations, as in stochastic dynamics (Yun-Yu et al., 1988, van Gunsteren, 1993). The main implicit influences of solvent, *i. e.* hydrophobic or structural effect, dielectric screening, random fluctuations and viscous drag, are mimicked by a modification of the interaction function (different functional form, additional terms, see *e. g.* Fraternali & van Gunsteren, 1996) and of the equations of motion (Langevin equation).

2.4 Low-resolution force fields

The purpose of low-resolution force fields is the study of large systems, while addressing long time-scale phenomena, such as fold recognition in proteins, protein folding, *de novo* protein design and drug design. With the currently available computing power, these problems are difficult to address, using force fields at atomic resolution (Hünenberger et al., 1995a,b, van

Gunsteren et al., 1995b). Force fields at the amino-acid residue level are being developed for peptides and proteins (Gerber, 1992, Jones, 1994, Ulrich et al., 1997). The main difficulty is to find an adequate expression for the interaction between residues that provides a sufficient energetical resolution to discriminate correct from incorrect structures. Once a functional form is selected, the interaction function parameters are usually calibrated via a statistical analysis of native protein structures. The effects of solvent are normally treated by a mean force term (section 2.3). A correct description of the dynamics is usually not expected from such models.

2.5 Hybrid force fields

A whole variety of models include the combination of a treatment of a few degrees of freedom at a high particle resolution and a treatment of the others at a lower resolution. For instance, the first or first few hydration shells of a macromolecule may be included explicitly in a simulation, the bulk solvent being modelled through a mean force (section 2.3). Another typical example is the simulation of chemical, or acid- or base-catalyzed reactions, in solution or in enzymes (Warshel, 1991, Field, 1993, Whitnell & Wilson, 1993, Liu et al., 1994, 1996a,b). Clearly, a quantum mechanical description of the electrons or the protons is required. However, due to the computational costs, such a treatment cannot be applied to the full system under study, and only a few relevant degrees of freedom can be treated in this way. Finding the proper interface between the different degrees of particle resolution in such hybrid models is here the main difficulty.

3 Force field terms

3.1 Expression of the classical Hamiltonian

As in the quantum description of a molecular system, the classical Hamiltonian (total energy of the system) depends simultaneously on the coordinates and the momenta of all particles in the system. In a similar manner as in Hartree-Fock calculations, where the quantum mechanical Hamiltonian is approximated by a sum of one and two electron operators, the classical Hamiltonian can be approximated by a sum of n-body terms

$$\mathcal{H}_{class}\left(\{\vec{q}_i, \vec{p}_i\}\right) \approx \sum_i \left[{}^{(1)}K(\vec{p}_i) + {}^{(1)}V(\vec{q}_i) \right] + \sum_i \sum_{j>i} {}^{(2)}V(\vec{q}_i, \vec{q}_j) +$$

$$\sum_i \sum_{j>i} \sum_{k>j} {}^{(3)}V(\vec{q}_i, \vec{q}_j, \vec{q}_k) + \cdots \tag{1}$$

where $i, j, k, \ldots$ are indices running over the N particles constituting the system, or a subset of these, $\vec{q}_i$ and $\vec{p}_i$ are the coordinate and momentum vectors of particle i, and the (n) superscripts indicate the order of the terms. The three (single or multiple) sums in Equation (1) correspond to the first three n-body terms of a force-field $i.\,e.$ $n = 1, 2, 3$. The principal terms that are used in current force fields, either with a physical or a non-physical ($i.\,e.$ ad hoc, to perturb the system or impose restraints derived from experimental information) meaning, are listed in Table II. The computational effort for calculating a n-body interaction term is either (i) of the order $\mathcal{O}(M)$, M being the length of a list of possible combinations of indices entering the multiple sums of (1), if such a list is available, or (ii) of the order $\mathcal{O}\left(\frac{N!}{(n!(N-n)!)}\right)$, N being the number of atoms in the system, if all combinations have to be calculated. Covalent interactions

are typically of type (i), whereas non-bonded interactions are of type (ii). For systems of a reasonable size, N^2 will always be larger than M for any list of covalent interactions, and the bulk of computer time will be used to calculate two-body non-bonded interactions. The computation of $\mathcal{H}_{class}$ is thus essentially an $\mathcal{O}(N^2)$ problem. Even for relatively small systems, the inclusion of three-body non-bonded terms is extremely expensive (Curtiss & Jurgens, 1990, Elrod & Saykally, 1994). On the other hand, the evaluation of a single N-body term is an inexpensive problem. An example may be the radius of gyration interaction that can be used to force protein unfolding in a molecular dynamics simulation (Hünenberger et al., 1995a).

3.2 Bond-stretching term

3.2.1 Functional forms

When simulations are performed at room temperature, and when no chemical (bond breaking) reaction is involved, bond lengths usually remain close to their equilibrium values. The bond stretching contribution to the potential energy can then be approximated adequately by a Taylor series (Hagler & Ewig, 1994)

$$E_b\left(\{b_i\}; \{b_i^0, {}^{(2)}k_{b,i}, {}^{(3)}k_{b,i}\}\right) = \sum_{\text{all bonds } i} \left[{}^{(2)}k_{b,i}(b_i^0 - b_i)^2 + {}^{(3)}k_{b,i}(b_i^0 - b_i)^3 + \cdots\right] \qquad (2)$$

where b^0 is the equilibrium bond length and ${}^{(n)}k_b$ the force constant corresponding to the term of power n. There is no first order term since the derivative of the potential energy has to be zero when $b = b^0$. For example, in the MM2 force-field (Allinger, 1977, Bowen & Allinger, 1991), terms are retained till the third (cubic) power. This has the disadvantage that the potential becomes negative for high internuclear separation and thus, an inadequate coordinate choice may cause bond dissociation. A quartic expansion is used in the MM3 (Allinger et al., 1989, Bowen & Allinger, 1991) and CFF93 (Maple et al., 1994a,b, Hwang et al., 1994) force fields, which fixes this problem. Although the inclusion of anharmonic terms ($n > 2$) clearly improves the description of vibrational properties of molecules in the gas phase, it may not do so in other applications. When oscillations with large amplitudes are considered, when the effect of non-bonded strain on a bond length and stretching frequency are of interest, or when the breaking of a bond is required, other potential forms can be used. For example, as in the CVFF force field (Hagler et al., 1979a–c, Lifson et al., 1979), a Morse-type function may be used

$$E_{Morse}\left(\{b_i\}; \{b_i^0, D_i, \alpha_i\}\right) = \sum_{\text{all bonds } i} D_i \left[e^{\alpha_i(b_i^0 - b_i)} - 1\right]^2 \qquad (3)$$

where D is the well depth, b^0 the equilibrium bond length and α a unitless parameter determining the width of the well. This equation already encompasses anharmonicities and provides a better description than a limited Taylor expansion around and away from the equilibrium bond length. Other possible three-parameter functions (Zavitsas et al., 1989, Ermler & Hsieh, 1990) are the Durham, Linnett, Lippincott, Rydberg, Simons-Parr-Finlan and Varshni functions. Most of these have been calculated *a priori* or tailored for diatomic molecules, but at least some may be applied successfully to individual bonds in polyatomic molecules (Brown & Truhlar, 1985). It has also been proposed that Taylor expansions in $(b^0 - b)^{-1}$ may be more adequate than expansions in $(b^0 - b)$ for fitting bond stretching energies from *ab initio* calculations, and a good correspondence has been observed over a wide range of lengths

(Dinur & Hagler, 1994). The use of such dissociative functions for modelling a bond breaking process remains, however, limited to specific systems and chemical reactions because (i) they are difficult to parametrize, and (ii) in the general case, the effect of bond breaking is not only local to a single bond and implies corresponding changes in the parameters of other covalent and non-bonded interaction terms.

In a large number of applications (*e. g.* AMBER, CHARMM, GROMOS, ...) and especially for the simulation of large molecules or the simulation of systems in explicit solvent, the detailed formalisms mentioned above are not used. A Taylor expansion limited to the second order (harmonic) is assumed to be sufficient since (i) the high bond-stretching and bond-angle bending frequencies are weakly coupled to the rest of the system and (ii) the low frequency motions (conformational changes, solvent relaxation) largely determine the thermodynamic properties of the system. The evaluation of the bond-stretching interaction may be made less expensive by using the quartic expression

$$E_b\left(\{b_i\}; \{b_i^0, k_{b,i}\}\right) = \sum_{\text{all bonds } i} k_{b,i} \left[(b_i^0)^2 - (b_i)^2\right]^2 \tag{4}$$

which avoids a square root operation in the calculation of the force and energy. In molecular dynamics simulations, since a proper integration of the (uninteresting) high-frequency bond stretching vibrations requires time steps of the order of 0.5 fs, a further (and common) time-saving technique is to constrain the bonds to their equilibrium lengths, which allows for the use of time-steps $4-5$ times longer without substantially affecting the dynamics (van Gunsteren & Karplus, 1982). It has been shown, however, that the bond angles should not be constrained simultaneously.

3.2.2   Combination rules

Combination rules for covalent bond interaction parameters are usually given in the form of a table as a function of the atom types of the atoms that define the bond. An interesting exception is the DREIDING force-field (Mayo et al., 1990), which uses an arithmetic combination rule

$$b_i^0(a, b) = R^0(a) + R^0(b) - 0.01 \text{ Å} \tag{5}$$

where a and b are the atom types of atoms forming bond i, and $R^0(a)$, $R^0(b)$ are the covalent radii corresponding to these atom types. The (harmonic) bond-stretching force constant is determined solely by the bond order.

3.3   Bond-angle bending term

3.3.1   Functional forms

Most of the considerations applying to bond stretching terms also apply here. For small deformations around the equilibrium bond angle, a Taylor expansion can be used

$$E_\vartheta\left(\{\vartheta_i\}; \{\vartheta_i^0, {}^{(2)}k_{\vartheta,i}, {}^{(3)}k_{\vartheta,i}, \dots\}\right) = \sum_{\text{all angles } i} \left[{}^{(2)}k_{\vartheta,i}(\vartheta_i^0 - \vartheta_i)^2 + {}^{(3)}k_{\vartheta,i}(\vartheta_i^0 - \vartheta_i)^3 + \cdots\right] \tag{6}$$

where ϑ^0 is the equilibrium angle and ${}^{(n)}k_\vartheta$ the force constant corresponding to the term of power n. For example, an expansion up to the fourth power is used in CFF93 (Maple et al.,

1994a,b, Hwang et al., 1994), the second and sixth power terms are retained in MM2 (Allinger, 1977, Bowen & Allinger, 1991), and MM3 (Allinger et al., 1989, Bowen & Allinger, 1991) uses a full expansion up to the sixth power. An alternative potential energy term which is used in some force fields, such as the CHARMM all-atom force field for DNA (MacKerell Jr. et al., 1995) is the Urey-Bradley energy term

$$
E_\vartheta \left(\{\vartheta_i\}; \{\vartheta_i^0, k_{\vartheta,i}, {}^{(1)}k_{d,i}, {}^{(2)}k_{d,i}\} \right) =
$$

$$
\sum_{\text{all angles } i} \left[k_{\vartheta,i}(\vartheta_i^0 - \vartheta_i)^2 + {}^{(1)}k_{d,i}(d_i^0 - d_i) + {}^{(2)}k_{d,i}(d_i^0 - d_i)^2 \right] \tag{7}
$$

where d_i is the 1,3 distance between atoms forming the extremity of the angle, d_i^0 its equilibrium value and ${}^{(n)}k_d$ the force constant corresponding to the term of power n. It has been shown that if E is defined to within a constant and d_i^0 is replaced by an effective distance, the linear term in Equation (7) can be omitted without loss of information (Pettitt & Karplus, 1985). The Urey-Bradley function already includes some anharmonicity as well as a coupling between the angle and the constituting bonds.

Again, in a number of applications (*e.g.* AMBER, CHARMM, GROMOS, ...) dealing with large molecules or molecules in the bulk phase, only the harmonic term is retained in Equation (6). A harmonic function in the angle cosine is also sometimes used (Mayo et al., 1990) for computational efficiency

$$
E_\vartheta \left(\{\vartheta_i\}; \{\vartheta_i^0, k_{\vartheta,i}\} \right) = \sum_{\text{all angles } i} k_{\vartheta,i}(\cos \vartheta_i^0 - \cos \vartheta_i)^2 \tag{8}
$$

3.3.2   Combination rules

Combination rules for bond angle bending parameters are usually given in the form of a table as a function of the atom types of the atoms that define the angle. An algebraic empirical combining rule for estimating harmonic angle bending from *ab initio* results or spectroscopic force fields has however been proposed (Halgren, 1990)

$$
k_{\vartheta,i}(a, b, c) = K \, Z(a) \, C(b) \, Z(c) \, \left(b_{ab}^0 + b_{bc}^0 \right)^{-1} \left(\vartheta_{abc}^0 \right)^{-2} \exp\left(-2 \frac{\left(b_{ab}^0 - b_{bc}^0 \right)^2}{\left(b_{ab}^0 + b_{bc}^0 \right)^2} \right) \tag{9}
$$

where a, b and c are the atom types of atoms forming angle i, K a constant, $Z(a)$, $C(b)$ and $Z(c)$ parameters depending solely on the atom types, and ϑ^0, b^0 equilibrium parameters.

3.4   Torsional dihedral angle term

3.4.1   Functional forms

If small oscillations around an equilibrium conformation are considered, the torsional potential energy term can, just as the bond stretching and bond angle bending terms, be expanded in a Taylor series. In most applications, however, when the relative energies of different conformers and the corresponding isomerisation barriers are of interest, or when conformational transitions are studied by molecular dynamics, Taylor series cannot be used. In these cases, the torsional

angle potential energy term needs to be 2π-periodic and symmetric at 0 and π, and can be expressed in terms of a cosine series

$$E_\varphi\left(\{\varphi_i\}; \{^{(1)}k_{\varphi,i}, {}^{(2)}k_{\varphi,i}, {}^{(3)}k_{\varphi,i}, \dots\}\right) =$$

$$\sum_{\text{dihedrals } i} \left[{}^{(1)}k_{\varphi,i}(1 - \cos\varphi_i) + {}^{(2)}k_{\varphi,i}(1 - \cos 2\varphi_i) + {}^{(3)}k_{\varphi,i}(1 - \cos 3\varphi_i) + \cdots \right] \qquad (10)$$

where $^{(n)}k$ is the force constant corresponding to the term of order n. For example, CFF93 (Maple et al., 1994a,b, Hwang et al., 1994) and MM3 (Allinger et al., 1989, Bowen & Allinger, 1991) use the first three terms in the expansion. The terms (of order n) are sometimes formulated in a slightly different way (Brooks et al., 1983, van Gunsteren & Berendsen, 1987, 1990), such as

$$|^{(n)}k_{\varphi,i}| - {}^{(n)}k_{\varphi,i}\cos n\varphi_i$$

$$\text{or} \quad {}^{(n)}k'_{\varphi,i}\left(1 + \cos(n\varphi_i - {}^{(n)}\delta_i)\right) \qquad \text{with} \qquad {}^{(n)}k'_{\varphi,i} > 0 \quad \text{and} \quad {}^{(n)}\delta_i = 0, \pi \qquad (11)$$

where $^{(n)}\delta$ in the second formulation is a phase shift, which plays the same role as the sign of $^{(n)}k_\varphi$ in the first formulation. Since the slope of the potential has to vanish at 0 and π, the only possible values of $^{(n)}\delta$ are 0 and π. If $^{(n)}k_\varphi$ is negative or $^{(n)}\delta_i$ is 0, the term has a maximum for $\varphi = 0$. If $^{(n)}k_\varphi$ is positive or $^{(n)}\delta_i$ is π, it has a minimum for $\varphi = 0$. These two latter formulations ensure that the potential is zero at the minimum of the curve, which may not be true for Equation (10). The number of terms retained in the cosine expansion varies from one force field to another and from one dihedral type to another. One should also note that the summation in Equation (10) need not include all dihedral angles, but may comprise only one or two particular dihedral angles out of the one to nine dihedrals that can be defined around a central covalent bond between two atoms having up to four covalently bound neighbour atoms.

3.4.2 Combination rules

Combination rules for torsional potential parameters are usually given in the form of a table as a function of the atom types of the four atoms, or of the two central atoms, that define the torsional angle.

3.5 Out-of-plane coordinate distortion term

3.5.1 Functional forms

In principle, the valence terms of a force field could be entirely defined in terms of bond lengths, bond angles and torsional dihedrals, as is for instance the case in the alkane CFF93 force field (Maple et al., 1994a,b, Hwang et al., 1994). There are however two reasons for introducing out-of-plane coordinate potential energy terms: (i) all the covalent internal coordinates mentioned till now can be expressed in terms of scalar products of vectors and there is thus no term to enforce chirality (which is just determined by the coordinates and sufficiently high isomerisation barriers). Enforcing the geometry around a site by using six bond angles (tetrahedral case) or three bond angles (planar case) without including cross-terms leads to an unrealistically stiff energy function (e. g. around a carbonyl carbon, the function tends to become quartic for out of plane bending) and (ii) when tetrahedral united atom carbons, CHR_3, are used, one of the

hydrogen is not explicitly present for the definition of an angle, but pyramidal inversion need be avoided.

The out-of-plane coordinate energy term should describe how difficult it is to force a non-planar geometry (trigonal site) or a non-tetrahedral geometry (tetracoordinated site or CHR_3 united atom). The functional form is most often chosen to be harmonic

$$E_\xi\left(\{\xi_i\}; \{\xi_i^0, k_{\xi,i}\}\right) = \sum_{\substack{\text{out-of-plane} \\ \text{coordinates } i}} k_{\xi,i}(\xi_i^0 - \xi_i)^2 \tag{12}$$

The definition of the out-of-plane coordinate is not unique (Dinur & Hagler, 1991). It can either be expressed in terms of (i) an improper dihedral angle, $i.\,e.$ the dihedral angle defined by a bond from the central atom to a peripheral atom, the vector from this peripheral atom to a second peripheral atom, and the vector from this second peripheral atom to a third peripheral atom, (ii) the height of the central atom above the plane defined by the three peripheral ones and (iii) the angle between one bond to the central atom and the plane defined by the central atom and the two peripheral atoms not involved in this bond. The summation in Equation (12) runs over a specified set of out-of-plane coordinates.

3.5.2   Combination rules

Combination rules for out-of-plane coordinate potential energy parameters are usually given in the form of a table as a function of the atom types of the four atoms, or of the two outer atoms, that define the coordinate.

3.6   Valence coordinates cross terms

3.6.1   Functional forms

It has been shown that the inclusion of valence coordinates coupling terms (off-diagonal terms) significantly improves the capacity of an empirical function to reproduce trends in the energy, and its first and second derivatives with respect to the atomic coordinates, from *ab initio* molecular orbital calculations (Halgren, 1990, Maple et al., 1994a) and trends from experimental data in vacuum (Hwang et al., 1994). The MM2 (Allinger, 1977, Bowen & Allinger, 1991) and CVFF (Hagler et al., 1979a-c, Lifson et al., 1979) force fields contain some of these terms, whereas the MM3 (Allinger et al., 1989, Bowen & Allinger, 1991) and CFF93 (Maple et al., 1994a) force fields use them in a systematic way. The commonly included off-diagonal terms are listed below (see also Table II). Interpretation of the terms in terms of force constants is given here with respect to the CFF93 force field for alkanes.

A. Bond – bond coupling (bonds j sharing one common atom with bond i)

$$E_{bb'}\left(\{b_i, b_j\}; \{b_i^0, b_j^0, k_{bb',ij}\}\right) = \sum_{\text{bonds } i} \sum_{\text{bonds } j > i}^{(\leq 6)} k_{bb',ij}(b_i^0 - b_i)(b_j^0 - b_j) \tag{13}$$

This term is present in CVFF and CFF93. Since k is positive, this term favours asymmetric bond stretching around a given site.

B. Bond-angle – bond coupling (two bonds j involved in the angle i)

$$E_{\vartheta b}\Big(\{\vartheta_i, b_j\}; \{\vartheta_i^0, b_j^0, k_{\vartheta b, ij}\}\Big) = \sum_{\text{angles } i} \sum_{\text{bonds } j}^{(2)} k_{\vartheta b, ij}(\vartheta_i^0 - \vartheta_i)(b_j^0 - b_j) \qquad (14)$$

This term is present in CVFF, CFF93, MM2 and MM3. It is used to reproduce vibrational frequencies and the bond-length effects in strained molecules where a bond angle is stretched or compressed. Since k is positive, bond lengthening is favoured when the bond angle is reduced.

C. Bond-angle – bond-angle coupling (angles j sharing one common bond with angle i)

$$E_{\vartheta \vartheta'}\Big(\{\vartheta_i, \vartheta_j\}; \{\vartheta_i^0, \vartheta_j^0, k_{\vartheta \vartheta', ij}\}\Big) = \sum_{\text{angles } i} \sum_{\text{angles } j}^{(\leq 10)} k_{\vartheta \vartheta', ij}(\vartheta_i^0 - \vartheta_i)(\vartheta_j^0 - \vartheta_j) \qquad (15)$$

This term is present in CVFF, CFF93 and MM3. It is used to reproduce vibrational frequencies for coupled bending modes. k may be positive or negative.

D. Torsional-angle – bond coupling (central bond or two peripheral bonds j involved in torsion i)

$$E_{\varphi b}\Big(\{\varphi_i, b_j\}; \{b_j^0, {}^{(1)}k_{\varphi b, i}, {}^{(2)}k_{\varphi b, i}, {}^{(3)}k_{\varphi b, i}, \dots\}\Big) =$$

$$\sum_{\text{dihedrals } i} \sum_{\text{bonds } j}^{(1) \text{ or } (2)} \Big(b_j^0 - b_j\Big) \Big[{}^{(1)}k_{\varphi b, i} \cos \varphi_i + {}^{(2)}k_{\varphi b, i} \cos 2\varphi_i + {}^{(3)}k_{\varphi b, i} \cos 3\varphi_i + \cdots\Big] \qquad (16)$$

This term is present in CFF93 up to order three and MM3 (torsion-central bond coupling only). It is used for reproducing the structures of molecules in which different conformers exhibit significant differences in bond lengths and bond angles. Since ${}^{(1)}k_\varphi$ is positive for the coupling to the central bond, a lenghtening of this bond is favoured in the eclipsed conformations. For peripheral bonds, k is negative and small.

E. Torsional-angle – bond-angle coupling (two angles j involved in torsion i)

$$E_{\varphi \vartheta}\Big(\{\varphi_i, \vartheta_j\}; \{\vartheta_j^0, {}^{(1)}k_{\varphi \vartheta, i}, {}^{(2)}k_{\varphi \vartheta, i}, {}^{(3)}k_{\varphi \vartheta, i}, \dots\}\Big) =$$

$$\sum_{\text{dihedrals } i} \sum_{\text{angles } j}^{(2)} \Big(\vartheta_j^0 - \vartheta_j\Big) \Big[{}^{(1)}k_{\varphi \vartheta, i} \cos \varphi_i + {}^{(2)}k_{\varphi \vartheta, i} \cos 2\varphi_i + {}^{(3)}k_{\varphi \vartheta, i} \cos 3\varphi_i + \cdots\Big] \qquad (17)$$

This term is present in CFF93 up to order three and plays a similar role as the term under D.

F. Torsional-angle – bond-angle – bond-angle coupling (two angles j and k involved in the torsion i)

$$E_{\varphi \vartheta \vartheta'}\Big(\{\varphi_i, \vartheta_j, \vartheta_k\}; \{\vartheta_j^0, \vartheta_k^0, k_{\varphi \vartheta \vartheta', i}\}\Big) = \sum_{\text{dihedrals } i} k_{\varphi \vartheta \vartheta', i}(\vartheta_j^0 - \vartheta_j)(\vartheta_k^0 - \vartheta_k) \cos \varphi_i \qquad (18)$$

This term is present in CVFF and CFF93. Since k is negative, an increase in the bond angles is favoured in the eclipsed conformations.

Note that the inclusion of 1,4 non-bonded interactions (atoms separated by three bonds) implicitly includes terms of type C – F. The force-constants in Equations (13) – (18) are obtained by fitting to the energy and its first and second derivative respective to the coordinates, calculated by *ab initio* techniques using a set of distorted structures of the molecule (Maple et al., 1994a). They can later be scaled using a limited number of parameters in order to reproduce experimental data, the assumption being that the errors made in the *ab initio* calculation are systematic (Hwang et al., 1994). Such type of calculations has only currently been performed systematically for alkanes.

3.7 Van der Waals interaction

3.7.1 Functional form

It is usually assumed that the non-electrostatic component of the interaction between non-bonded atoms can be described in the same way as the interaction between rare gas atoms, *i. e.* a long-range weak attraction due to induced-dipole induced-dipole (dispersion) interaction and a short-range steep repulsion due to the overlap of the electron clouds. This type of interaction is given the generic name of van der Waals interaction. Although the features mentioned above are generally accepted, the proper functional description of van der Waals interactions is, however, still matter of discussion (Halgren, 1992, Hart & Rappe, 1992a,b). Most current force fields use a 12-6 van der Waals function (Lennard-Jones function), where the steep repulsion is described by a $1/r_{ij}^{12}$ dependence and the dispersion by a $1/r_{ij}^{6}$ dependence. Three equivalent definitions can be found in the literature

$$
E_{12-6}\left(\{r_{ij}\}; \{C_{12}(i,j), C_6(i,j)\}\right) =
$$
$$
\sum_{i}^{N_{atoms}} \sum_{j>i}^{N_{atoms}} \left[\frac{C_{12}(i,j)}{r_{ij}^{12}} - \frac{C_6(i,j)}{r_{ij}^{6}} \right]
$$

or
$$
E_{12-6}\left(\{r_{ij}\}; \{R_{min}(i,j), \varepsilon(i,j)\}\right) =
$$
$$
\sum_{i}^{N_{atoms}} \sum_{j>i}^{N_{atoms}} \varepsilon(i,j) \left[\left(\frac{R_{min}(i,j)}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{min}(i,j)}{r_{ij}} \right)^{6} \right] \qquad (19)
$$

or
$$
E_{12-6}\left(\{r_{ij}\}; \{\sigma(i,j), \varepsilon(i,j)\}\right) =
$$
$$
\sum_{i}^{N_{atoms}} \sum_{j>i}^{N_{atoms}} 4\varepsilon(i,j) \left[\left(\frac{\sigma(i,j)}{r_{ij}} \right)^{12} - \left(\frac{\sigma(i,j)}{r_{ij}} \right)^{6} \right]
$$

In Equation (19), R_{min} is the distance at minimum energy, σ the distance at zero energy ($\sigma < R_{min}$) and ε the well depth at the minimum (with respect to infinite separation). These three parameters keep their meaning even if the energy is not described by a 12-6 van der Waals function. Since the experimental energy functions for rare gases follow a single reduced form around the minimum (Kestin et al., 1984, Halgren, 1992), R_{min} or σ definitions and combination rules are interchangeable to a large extent. In the case of a 12-6 van der Waals function, the conversion between the definitions of Equation (19) is straightforward

$$R_{min}(i,j) = \sqrt[6]{\frac{2C_{12}(i,j)}{C_6(i,j)}}$$

$$\sigma(i,j) = \sqrt[6]{\frac{C_{12}(i,j)}{C_6(i,j)}} \tag{20}$$

$$\varepsilon(i,j) = \frac{C_6^2(i,j)}{4C_{12}(i,j)}$$

It has been suggested that a softer van der Waals interaction might perform better than a 12-6 form. For example, a 9-6 van der Waals interaction is used in CVFF (Hagler et al., 1979a–c, Lifson et al., 1979) or CFF93 (Maple et al., 1994a)

$$E_{9-6}\Big(\{r_{ij}\}; \{R_{min}(i,j), \varepsilon(i,j)\}\Big) =$$

$$\sum_i^{N_{atoms}} \sum_{j>i}^{N_{atoms}} \varepsilon(i,j) \left[2\left(\frac{R_{min}(i,j)}{r_{ij}}\right)^9 - 3\left(\frac{R_{min}(i,j)}{r_{ij}}\right)^6 \right] \tag{21}$$

Another widely used function is the exp-6 function (Mayo et al., 1990), which reads

$$E_{exp-6}\Big(\{r_{ij}\}; \{\varepsilon(i,j), R_{min}(i,j), \zeta(i,j)\}\Big) =$$

$$\sum_i^{N_{atoms}} \sum_{j>i}^{N_{atoms}} \frac{\varepsilon(i,j)}{\zeta(i,j) - 6} \left[6\exp\left(\zeta(i,j)\left(1 - \frac{r_{ij}}{R_{min}(i,j)}\right)\right) - \zeta(i,j)\left(\frac{r_{ij}}{R_{min}(i,j)}\right)^{-6} \right] \tag{22}$$

where ζ is a dimensionless scaling parameter. Although all types of functions defined above perform similarly close to the equilibrium distances, it has been suggested both using theoretical arguments and comparison to *ab initio* results (Hart & Rapé, 1992a,b) that a Morse-type function may perform better over a wider range of distances

$$E_{Morse}\Big(\{r_{ij}\}; \{\varepsilon(i,j), R_{min}(i,j), \alpha(i,j)\}\Big) =$$

$$\sum_i^{N_{atoms}} \sum_{j>i}^{N_{atoms}} \varepsilon(i,j) \left[\exp\left[2\alpha(i,j)\left(1 - \frac{r_{ij}}{R_{min}(i,j)}\right)\right] \right.$$

$$\left. - 2\exp\left[\alpha(i,j)\left(1 - \frac{r_{ij}}{R_{min}(i,j)}\right)\right] \right] \tag{23}$$

where α is a dimensionless scaling parameter.

Finally, a buffered 14-7 energy function has also been proposed (Halgren, 1992)

$$E_{buf\text{-}n\text{-}m}\Big(\{r_{ij}\}; \delta, \gamma, \{\varepsilon(i,j), R_{min}(i,j)\}\Big) =$$

$$\sum_i^{N_{atoms}} \sum_{j>i}^{N_{atoms}} \varepsilon(i,j) \left[\frac{(1+\delta)R_{min}(i,j)}{r_{ij} + \delta R_{min}(i,j)}\right]^{n-m} \left[\frac{(1+\gamma)R_{min}^m(i,j)}{r_{ij}^m + \gamma R_{min}^m(i,j)} - 2\right] \tag{24}$$

with $n = 14$, $m = 7$, $\delta = 0.07$ and $\gamma = 0.12$, these parameters being obtained from a best fit to rare gas experimental data. A correct description of van der Waals interactions over a broad range of distances is essential for condensed phase simulation, since a broad spectrum of interatomic distances will be present.

3.7.2 Combination rules

Since the definition of N atom types implies the definition of $\frac{1}{2}N(N+1)$ van der Waals interaction parameter sets for atom pairs, most force-fields make use of combination rules which depend on sets of N atomic parameters, which can be calibrated by studying the homonuclear case (Waldmann & Hagler, 1993, Halgren, 1992). For Equation (19), (21)–(24), many types of combination rules are found in the literature:

A. Geometric means (van Gunsteren, 1987)

The following rules are equivalent for the case of a 12-6 van der Waals interaction (the latter two for any van der Waals interaction)

$$
\begin{aligned}
& C_6(i,j) = \sqrt{C_6(i,i)\,C_6(j,j)} && \text{and} && C_{12}(i,j) = \sqrt{C_{12}(i,i)\,C_{12}(j,j)} \\
\text{or}\quad & R_{min}(i,j) = \sqrt{R_{min}(i,i)\,R_{min}(j,j)} && \text{and} && \varepsilon(i,j) = \sqrt{\varepsilon(i,i)\,\varepsilon(j,j)} \\
\text{or}\quad & \sigma(i,j) = \sqrt{\sigma(i,i)\,\sigma(j,j)} && \text{and} && \varepsilon(i,j) = \sqrt{\varepsilon(i,i)\,\varepsilon(j,j)}
\end{aligned}
\tag{25}
$$

B. Geometric mean for ε and arithmetic mean for R_{min} or σ (Lorentz-Berthelot mixing rules)

The following rules are equivalent for the case of any n–m van der Waals interaction

$$
\begin{aligned}
& R_{min}(i,j) = \tfrac{1}{2}\big[R_{min}(i,i) + R_{min}(j,j)\big] && \text{and} && \varepsilon(i,j) = \sqrt{\varepsilon(i,i)\,\varepsilon(j,j)} \\
\text{or}\quad & \sigma(i,j) = \tfrac{1}{2}\big[\sigma(i,i) + \sigma(j,j)\big] && \text{and} && \varepsilon(i,j) = \sqrt{\varepsilon(i,i)\,\varepsilon(j,j)}
\end{aligned}
\tag{26}
$$

C. Arithmetic mean for R_{min}^6 and geometric mean for εR_{min}^6

This combination rule has been proposed recently and tested for rare gases (Waldmann & Hagler, 1993)

$$
R_{min}(i,j) = \sqrt[6]{\frac{R_{min}^6(i,i) + R_{min}^6(j,j)}{2}}
$$

$$
\varepsilon(i,j) = \frac{1}{R_{min}^6(i,j)}\sqrt{\varepsilon(i,i)R_{min}^6(i,i)\,\varepsilon(j,j)R_{min}^6(j,j)}
\tag{27}
$$

D. Cubic-mean rule for R_{min} and HHG mean for ε

This combination rule has been proposed recently and tested for rare gases (Halgren, 1992), where the HHG mean is the harmonic mean of harmonic and geometric means

$$
R_{min}(i,j) = \frac{R_{min}(i,i)^3 + R_{min}(j,j)^3}{R_{min}(i,i)^2 + R_{min}(j,j)^2}
\tag{28}
$$

$$
\varepsilon(i,j) = \frac{4\varepsilon(i,i)\varepsilon(j,j)}{\left(\sqrt{\varepsilon(i,i)} + \sqrt{\varepsilon(j,j)}\right)^2}
\tag{29}
$$

E. Slater-Kirkwood combination

The Slater-Kirkwood expression (van Gunsteren & Karplus, 1982, Brooks et al., 1983, Halgren, 1992) is more than a combination rule, since it also allows the estimation of van der Waals parameters from experiment

$$C_6(i,j) = K \frac{\alpha(i)\alpha(j)}{\sqrt{\dfrac{\alpha(i)}{N(i)}} + \sqrt{\dfrac{\alpha(j)}{N(j)}}}$$

$$C_{12}(i,j) = \tfrac{1}{2}\, C_6(i,j)\left(R(i) + R(j)\right)^6 \tag{30}$$

where K is a constant, $\alpha(i)$ is the polarizability of atom i, $R(i)$ its van der Waals radius, and $N(i)$ its effective number of outer shell electrons.

Other combinations have been proposed, which involve additional parameters such as polarizability, ionization potentials or dispersion force coefficients. These are however not well suited for general empirical force fields, since one would like to restrict the number of parameters involved.

3.8 Electrostatic interactions

3.8.1 Functional forms

The correct treatment of electrostatic interactions is an essential but difficult problem in the design of empirical energy functions (Harvey, 1989, Davis & McCammon, 1990, van Gunsteren & Berendsen, 1990, Berendsen, 1993, Smith & van Gunsteren, 1993). This is mainly due to their long-range nature, which causes dependence on the system size and boundary conditions, as well as high computational costs. In condensed-phase simulations, these high computational costs, together with the use of periodic boundary conditions, require approximations, which will unfortunately influence the properties of the simulated system. In most cases, the interaction is defined in terms of a pairwise Coulomb interaction between point (atomic or virtual site) partial charges, that is, by a monopole approximation. The effect of the polarizability of the electron cloud is assumed to be included in the interaction between these point charges in an average manner, and these charges are thus effective charges. Ideally, the interaction should be calculated by scanning all charge pairs, *i. e.*

$$E_{Cb}\left(\{r_{ij}\}; \{q_i q_j\}\right) = \sum_{i}^{N} \sum_{j>i}^{N} \frac{1}{4\pi\epsilon_0\epsilon_1}\, \frac{q_i q_j}{r_{ij}} \tag{31}$$

where r_{ij} is the distance between charges i and j, $q_i q_j$ the product of the charges, ϵ_0 the permittivity of vacuum, ϵ_1 the relative permittivity of the medium and N the number of atoms in the system. Equation (30) is in principle exact, but practically directly applicable only to vacuum simulations of small isolated molecules, with the aim of reproducing vacuum properties. It cannot be used for:

A. Medium and large scale problems: since the computational expenses grow as N^2.

B. Fixed boundary problems: if the system consists of a molecule, plus possibly some layers of solvent, surrounded by vacuum, surface tension effects will distort its properties. In

the absence of dielectric screening from outside the system, the electrostatic interaction inside the system will be overestimated, and in the absence of van der Waals forces with the outside, the surface of the boundary will tend to become minimal (spherical shape). When explicit solvent molecules are present, evaporation may also occur.

C. Periodic boundary problems: if the system consists of an infinite series of replicas of a central cell (periodic boundary conditions, for crystal or solution simulations), the number of pairs in Equation (30) is infinite.

A wealth of techniques have been designed to find approximate treatments which remedy these problems, and try to find the best compromise between efficiency and accuracy. The following list is non-exhaustive:

1. *Boundary corrections (point B)*

The distortions induced at the interface to vacuum can be reduced by corrections which attempt to mimic the effect of solvent outside the boundary (King & Warshel, 1989, Beglov & Roux, 1994, Essex & Jorgensen, 1995, Wang & Hermans, 1995): (i) short range contacts, by addition of a soft-wall interaction or position restraining of the atoms in the surface layer, (ii) electrostatic effects at the boundary, by addition of dipole orientation interactions, and (iii) dynamical fluctuations, by the use of stochastic boundary condition. These boundary corrections are difficult to calibrate and often have to be reparametrized for each specific system considered.

2. *Redistribution and reduction of the charges (point B)*

The dielectric screening effect of a virtual solvent outside the boundary can be in first approximation, included by the use of a set of more distributed and reduced charges (van Gunsteren & Berendsen, 1987). This method is very *ad hoc*.

3. *Distance dependent dielectric (point B)*

The dielectric screening effect can also be mimicked by replacing ϵ_1 in Equation (30) by an effective dielectric constant ϵ_{eff}, proportional to the distance between charges, *i. e.* $\epsilon_{eff} = n \cdot r_{ij}$, usually with $n = 1, 4$ or 8 Å^{-1}. In this approximation, the screening effect is assumed proportional to the amount of bulk solvent between the charges, and thus, to the distance. This method is also *ad hoc* and lacks physical meaning.

4. *Screening functions (point B)*

The approach is similar to the previous one, but ϵ_1 is replaced by $\epsilon_{eff} e^{\kappa \, r_{ij}}$, where κ is the inverse Debye screening length. The choice of an adequate ϵ_{eff} (constant or function of r_{ij}) is problematic and the application to heterogeneous systems is not satisfactory.

5. *Continuum methods (point B)*

The system is assumed to be surrounded by a dielectric continuum of permittivity ϵ_2 (Wang & Hermans, 1995). The influence of the charge distribution in the system on the continuum outside the boundary, induces a reaction field potential inside the boundary. When the shape of the boundary is highly symmetric, the interaction can be computed analytically (Born, Onsager models). In other cases, it has to be computed numerically

(series expansion of the reaction field, finite difference, finite elements or boundary elements methods). Treatment of particles near the boundary is the major problem of these methods.

6. *Langevin dipoles (point B)*

The solvent is modelled by a set of polarizable and rotatable dipoles on a grid, which average orientation is described by a Langevin type equation (Russel & Warshel, 1985). The model is relatively inexpensive and seems more realistic than a continuum approximation. It is, however, difficult to parametrize, a proper description of the interface is problematic, and the properties of the system may depend on the grid parameters.

7. *Lattice sums (point C)*

These methods are based on an exact periodic treatment of the infinite system in simulations using periodic boundary condition (Allen & Tildesley, 1987, Smith & van Gunsteren, 1993, Fincham, 1994, Luty et al., 1995, Smith & Pettitt, 1995). The infinite sum over all atoms and periodic images in Equation (30) can be rewritten as two finite sums over lattice (real space) and reciprocal lattice (Fourier space) vectors, plus a constant self-energy term, which can in principle be computed exactly. These methods are, however, complicated to implement, sometimes computationally expensive and they may enforce long-range correlations through periodicity. These are realistic in simulations of crystals, but may give rise to artefacts in bulk-phase systems, although only under special circumstances (Figueirido et al., 1995, Luty & van Gunsteren, 1996). Lattice sum techniques include Ewald summation, particle-particle particle mesh and related methods.

8. *Minimum image convention (point C)*

The interaction is only calculated between charge i of the central cell and the closest periodic image of charge j. The number of pairs is then finite, but can become large, *i. e.* $\mathcal{O}(N^2)$. This convention is not much used, since all charges interacting with i belong to a volume of the same shape as the unit cell, which induces anisotropy effects.

9. *Simple spherical cut-off (points A,B,C)*

The long range correlation problems inherent to lattice sum methods (7) and the anisotropy problem inherent to the nearest image convention (8) can be reduced if the Coulomb interaction is zeroed at a given distance between charges, the cut-off distance R_c. The sphere of radius R_c (cut-off sphere) around a charge i has to be smaller than the unit cell, so that only nearest images are selected inside the cut-off. This method is simple to implement and allows for a significant reduction of the computational costs for large systems, since the effort is roughly $\mathcal{O}(NR_c^3)$, and R_c can be much smaller than the box volume for large systems. Although it is a good approximation for non-polar systems, it may however produce serious problems for polar systems (Neumann, 1983, Neumann et al., 1984), ionic systems (Brooks III et al., 1985, Brooks III, 1987, Madura & Pettitt, 1988) or biomolecules in solution (Schreiber & Steinhauser, 1992a,b,c), since the long range Coulomb force often differs significantly from zero at the cut-off distance. The main problems (Allen & Tildesley, 1987) are non-conservation of the energy for a microcanonical simulation, heating effects at the cut-off due to a non-zero force and structural, statistical and dielectric distortions over the whole range of intermolecular distances. The following points 9.1 – 9.6 describe possible corrections to the simple

spherical cut-off approximation, which attempt to minimize these distortions (Brooks III et al., 1985, Loncharich & Brooks, 1989, Steinbach & Brooks, 1994).

9.1 Charge-group interaction

Charges are grouped in terms of chemically (or intuitively) based charge groups, either neutral or bearing an integer (electronic) charge (*e. g.* carbonyl groups, amino groups . . .). The atom-based truncation is then replaced by a charge-group based cutoff criterion (van Gunsteren & Berendsen, 1987,1990). For two neutral charge groups I and J, the leading term in the electrostatic interaction takes a r_{IJ}^{-3} dependence, which reduces significantly the effects of truncation. They are, however, not completely eliminated (Neumann, 1983, Neumann et al., 1984, Brooks III, 1987, Madura & Pettitt, 1988). The inconvenience of the method is that it may require a modification of the original charge distribution as obtained, for example, from quantum-mechanical calculations.

9.2 Twin-range method

In this method, a second (long-range) cut-off R_L is introduced. Interaction of charge pairs with $R_C < r_{ij} < R_L$ are calculated every n time steps ($n > 1$, usually $\approx 5\text{--}10$, together with the pair list update) and assumed constant in-between (van Gunsteren & Berendsen, 1990). If the high frequency fluctuations in the long range forces are negligible, the effective cut-off is increased to R_L without significant additional computational costs. In a variant, the interaction between charge groups at distances between R_C and R_L are approximated by a multipole expansion, *e. g.* up to quadrupole interactions (Brooks et al., 1983)

9.3 Switching function

To avoid abrupt truncation of the interaction at the cut-off radius R_C, the Coulomb interaction can be multiplied by a so-called switching function, $SW(r_{ij}, R_S, R_C)$ with $R_S < R_C$, a continuous function with continuous derivative, which has the value 1 if $r_{ij} < R_S$ and 0 if $r_{ij} > R_C$ (Brooks et al., 1983, Loncharich & Brooks, 1989). Energy conservation is improved, heating effects are reduced, but structural artefacts are still observed.

9.4 Shifting function

Alternatively, the Coulomb interaction can be multiplied by a so-called shifting function, $SF(r_{ij}, R_C)$, a continuous function with continuous derivative, which has the value 1 at $r_{ij} = 0$ and 0 if $r_{ij} > R_C$ (Brooks et al., 1983, Prevost et al., 1990). The inconvenience of this method is that the interaction is changed over the whole range of r_{ij} distances from 0 to R_C.

9.5 Reaction field correction

The medium outside the spherical cut-off cavity may be approximated by a dielectric continuum of relative dielectric permittivity equal to that of the bulk solvent, ϵ_2 (Barker & Watts, 1973, Hummer et al., 1992, Barker, 1994, Chipot et al., 1994a,b, Wood, 1995). The influence of the charge distribution (limited to a dipolar term) inside the cut-off on the continuum outside induces a reaction field potential inside the cut-off sphere. This additional interaction can be described as a pairwise

interaction and added to the Coulomb term to give

$$E_{Cb+RF}\Big(\{r_{ij}\};\{q_iq_j\},R_{RF},\epsilon_2\Big) =$$

$$\sum_i^{N_{atoms}} \sum_{j>i}^{N_{atoms}} \frac{q_iq_j}{4\pi\epsilon_0\epsilon_1}\left[\frac{1}{r_{ij}} + \frac{\epsilon_2-\epsilon_1}{2\epsilon_2+\epsilon_1}\frac{r_{ij}^2}{R_{RF}^3} - \frac{3\epsilon_2}{2\epsilon_2+\epsilon_1}\frac{1}{R_{RF}}\right] \tag{32}$$

where R_{RF} is in principle equal to R_C. When ϵ_2 is large, both the energy and the force vanish at the cut-off, and thus, this correction can be considered as a physically based shifting function. It makes a considerable difference whether this additional interaction is included during the simulation or as a correction afterwards (Daura et al., 1996). This formalism can be extended to include the effect of non-zero ionic strengths (Tironi et al., 1995). This treatment is a significant improvement to straight truncation, but not entirely correct when applied to heterogeneous systems. Its use might also require a force-field reparametrization (Smith & van Gunsteren, 1995).

9.6 Distance dependent dielectric

Since the approximation described under point (3) gives to the electrostatic energy a r_{ij}^{-2} dependence, and thus reduces truncation artefacts, it has also been used in condensed phase simulations under periodic boundary conditions. However there, the solvent is explicitly present, which means a double-counting of dielectric screening effects.

3.8.2 Combination rules

Formally, the Coulomb law has the form of a combination rule, since the magnitude of the interaction is proportional to the product of the charges of the individual atoms. Some force fields make further use of combination rules to determine the atomic point charges. In the bond increment method (Maple et al., 1994a, Oie et al., 1981), the charge of an atom i is calculated as

$$q_i = \sum_{\substack{\text{first neigh-}\\\text{bours } j}}^{(4)} \delta\Big(a(i),b(j)\Big) \tag{33}$$

where $a(i)$ and $b(j)$ are the atom types of i and j, respectively, and the function δ satisfies $\delta(a,b) = -\delta(b,a)$. This has the advantage that a single bond parameter is required to evaluate all charges, and that electroneutrality is always preserved.

3.9 Hydrogen-bonding term

An explicit hydrogen bonding interaction term is sometimes added to the already present non-bonded interactions described above. Its purpose is to avoid too short hydrogen bonds due to a strong electrostatic attraction, and to allow for a specific fine tuning of hydrogen-bond distances and energies. In some force fields, the van der Waals 12-6 parameters for hydrogen-bonded atoms are reduced at the same time. For example, in CHARMM (Brooks et al., 1983),

the hydrogen-bond potential energy is described by a sum of four-body terms

$$E_{hb}\Big(\{r_{AD}, \angle(A\cdots H\!-\!D), \angle(AA\!-\!A\cdots H)\}; \{C_\gamma, C_\delta, \gamma, \delta, m, n\}\Big) =$$

$$\sum_{\substack{\text{H-bonds} \\ AA-A\cdots H-D}} \left(\frac{C_\gamma}{r_{AD}^\gamma} - \frac{C_\delta}{r_{AD}^\delta}\right) \cos^m \angle(A\cdots H\!-\!D) \ \cos^n \angle(AA\!-\!A\cdots H) \tag{34}$$

where AA, A, H and D are the acceptor-antecedent, the acceptor, the hydrogen and the donor heavy atom, m depends on the type of D ($m = 0$, 2 or 4) and n on the type of A ($n = 0$ or 2). In CHARMM, the $\cos^m$ function is zeroed if its argument is less than 90° and the $\cos^n$ function if its argument is less than 90° and $n > 0$. Normally a 12-10 function is used for the radial dependence, $i.\,e.$ $\gamma = 12$ and $\delta = 10$. In other force fields ($e.\,g.$ Weiner et al., 1984), only the radial dependence is retained and a two-body 12-10 function is used ($i.\,e.$ $m = n = 0$ in Equation (33)). The presence of such a specific hydrogen-bonding interaction term requires some additional bookkeeping. If the structure is rigid enough, a permanent list of hydrogen-bonded groups can be defined. This list can also be automatically updated at regular intervals. If one assumes that the radial 12-10 correction can equally well be modelled by a 12-6 correction, it can be incorporated into the normal van der Waals interaction terms, as is done in GROMOS (van Gunsteren & Berendsen, 1987, 1990). This requires the use of a special combination rule for 12-6 van der Waals parameters, namely

$$C_6(i,j) = \sqrt{c_6(i, t_{ij})\, c_6(j, t_{ij})} \qquad \text{and} \qquad C_{12}(i,j) = \sqrt{c_{12}(i, t_{ij})\, c_{12}(j, t_{ij})} \tag{35}$$

where t_{ij} determines if the interaction between i and j is polar or not. This method offers the advantage that no special bookkeeping is required for the hydrogen bonds. The inconvenience is that even if the orientation of the D-H group is not optimal for making a hydrogen bond to A, the special van der Waals parameters will still be used for the A/D interaction. When only a radial function is used, there is no explicit angle dependence, but the implicit dependence of the electrostatic and repulsive van der Waals non-bonded interaction upon the hydrogen-bond angle plays a similar role.

4 Conclusion

In the present text, some of the main issues with respect to empirical classical force fields have been briefly described. Focusing mainly on models at atomic resolution, the terms that are most commonly found in the interaction energy function have been listed.

Since for the study of many large problems, the use of molecular orbital methods is computationally untractable, there is considerable interest in developing such empirical force fields. Due to the constant increase of the power of computers, the problems that can be addressed by these techniques increase regularly in size, complexity and in terms of the volume of conformational space that can be sampled. This in turn, triggers further development of the force fields themselves. New functional forms are proposed, which allow for a better energetic resolution in force fields, and systematic procedures begin to emerge for the parametrization of these functions based both on theoretical and experimental data.

ELEMENTARY UNIT	PHASE	TYPE OF INTERACTION (OPERATOR / FUNCTION)	DEGREES OF FREEDOM AVERAGED OUT	REF.
electrons and nuclei:	*gas phase:*	*ab initio*, density functional:	none	(a)
		first principle quantum mechanical Hamiltonian, Born-Oppenheimer surface		
		semi-empirical:	none	(b)
		approximated Hamiltonian		
	explicit solvent:	*idem*, supermolecule methods	none	(c)
	implicit solvent:	*idem*, additional reaction field potential	solvent	(d)
(united-)atoms:	*gas phase:*			
all atoms		classical empirical interaction function	electronic	(e)
united atoms (aliphatic groups only)		*idem*	aliphatic H	(e)
united atoms (all CH_n groups)		*idem*	all H bound to C	(e)
united atoms (all)		*idem*	all H	(e)
idem	*explicit solvent:*	*idem* include explicit solvent terms	*idem*	(f)
idem	*implicit solvent:*	*idem* possible corrections in the functional form, parameters or by additional terms or in the equation of motion	solvent	(f)
atom groups as "bead(s)": e. g. amino-acids in proteins represented by one or a few beads	*implicit solvent:* (or crystal)	statistically based interaction function	side-chain	(g)
molecules: represented by a sphere, a rod or a disk	*liquid phase:* (or crystal)	average intermolecular interaction function	intramolecular	(h)

(n)	Subset	Type	Term
1	all atoms	P	Kinetic energy
	charged atoms	P	Interaction with an external electric field
	surface atoms	P	Stochastic/frictional force on a macromolecule
	listed or all atoms	U	Atomic positional restraining
2	all atom pairs	P	Pairwise non-bonded interaction (point charges, point charge/point dipole ..., van der Waals, solvent accessible surface area interaction)
	bonded atoms	P	Covalent bond
	H-bonded atoms	P	H–bonding potential (Acceptor–Donor)
	listed atom pairs	U	Distance restraining
3	all atom triples	P	Triple non-bonded interactions (expensive, seldomly used)
	atoms in bond angle	P	Covalent bond-angle bending
	pairs of bonds	P	Bond–bond cross-term
	bonds in angle	P	Bond–angle cross-term
4	atoms in dihedrals	P	Torsional interaction, improper dihedral interaction
	H-bonded atoms	P	H–bonding (Acceptor Antecedent, Acceptor, Hydrogen, Donor)
	pairs of angles	P	Angle–angle cross-term (around one centre)
	atoms in dihedral	P	Bond–dihedral cross-term (central bond), Angle–angle–torsion cross-term
	atoms in dihedral	U	J-value restraining
≥ 5	covalent neighbours	P	Other cross-terms among bond, angles and dihedrals
N	all atoms	P	Point polarizability
	all atoms	U	Radius of gyration unfolding force

Table II: n-body interaction terms found in common force fields

(n): order of the term, $i.\,e.$ the number of particles involved in the interaction term, possibly all atoms (N), **Subset**: subset of atoms for which the term is calculated, either from a list or all atoms (pairs, triples, respectively), **Type**: physical (P) or unphysical (U) term.

References

Allen, M.P. & Tildesley, D.J. (1987). Computer simulation of liquids, Oxford University Press, Oxford, pp 1-385

Allinger, N.L. (1977). Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. *J. Am. Chem. Soc.* **99**, 8127–8134

Allinger, N.L., Yuh, Y.H. & Lii, J.-H. (1989). Molecular mechanics. The MM3 force field for hydrocarbons. 1. *J. Am. Chem. Soc.* **111**, 8551–8566

Ángyán, J.G. (1992). Common theoretical framework for quantum chemical solvent effect theories. *J. Math. Phys.* **10**, 93–137

Barker, J.A. & Watts, R.O. (1973). Monte Carlo studies of the dielectric properties of water-like models. *Mol. Phys.* **26**, 789–792

Barker, J.A. (1994). Reaction field, screening, and long-range interactions in simulations of ionic and dipolar systems. *Mol. Phys.* **83**, 1057–1064

Beglov, D. & Roux, B. (1994). Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. *J. Chem. Phys.* **100**, 9050–9063

Berendsen, H.J.C (1993). Electrostatic interactions. In: Computer simulation of biomolecular systems, theoretical and experimental applications, Vol. II, van Gunsteren, W.F., Weiner, P.K. & Wilkinson, A.J., Eds., ESCOM Science Publishers, B.V., Leiden, The Netherlands, pp 161–181

Beutler, T.C., Mark, A.E., van Schaik, R.C., Gerber, P.R. & van Gunsteren, W.F. (1994). Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **222**, 529–539

Beutler, T.C. & van Gunsteren, W.F. (1994). Molecular dynamics free energy calculation in four dimensions. *J. Chem. Phys.* **101**, 1417–1422

Binder, K. (1992). Topics in applied physics, Vol. 71: The Monte Carlo method in condensed matter physics, Springer-Verlag, Berlin

Bowen, J.P. & Allinger, N.L. (1991). Molecular mechanics: The art and science of parametrization. In: Reviews in computational chemistry, Vol. II, Lipkowitz, K.B. & Boyd, D.B., Eds., VCH Publishers Inc., New York, pp 81–97

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. & Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217

Brooks III, C.L., Pettitt, B.M. & Karplus, M. (1985). Structural and energetic effects of truncating long ranged interactions in ionic and polar fluids. *J. Chem. Phys.* **83**, 5897–5908

Brooks III, C.L. (1987). The influence of long-range force truncation on the thermodynamics of aqueous ionic solutions. *J. Chem. Phys.* **86**, 5156–5162

Brooks III, C.L., Karplus, M., Pettitt, B.M. (1988). Proteins: A theoretical perspective of dynamics, structure and thermodynamics, Wiley series on advances in chemical physics, Vol LXXI, Prigogine, I. & Rice, S., Eds., John Wiley & Sons, New York, pp 1-259

Brown, F.B. & Truhlar, D.G. (1985). Dissociation potential for breaking a C–H bond in methane. *Chem. Phys. Lett.* **113**, 441–446

Chipot, C., Millot, C., Maigret, B. & Kollman, P.A. (1994a). Molecular dynamics free energy simulations: Influence of the truncation of long-range nonbonded electrostatic interactions on the free energy calculations of polar molecules. *J. Chem. Phys.* **101**, 7953–7962

Chipot, C., Millot, C., Maigret, B. & Kollman, P.A. (1994b). Molecular dynamics free energy perturbation calculations: Influence of nonbonded parameters on the free energy of hydration of charged and neutral species. *J. Phys. Chem.* **98**, 11362–11372

Clark, M., Cramer III, R.D. & van Opdenbosch, N. (1989). Validation of the general purpose Tripos 5.2 force field. *J. Comput. Chem.* **10**, 982–1012

Cramer, C.J. & Truhlar, D.G. (1992). An SCF solvation model for the hydrophobic effect and absolute free energies of aqueous solvation. *Science* **256**, 213–217

Cramer, C.J. & Truhlar, D.G. (1994). Continuum solvation models: Classical and quantum mechanical implementations. In: Reviews in computational chemistry, Vol. VI, Lipkowitz, K.B. & Boyd, D.B., Eds., VCH Publishers Inc., New York, pp 1-70

Curtiss, L.A. & Jurgens, R. (1990). Nonadditivity of interaction in hydrated Cu^+ and Cu^{2+} clusters. *J. Am. Chem. Soc.* **94**, 5509–5513

Daura, X., Hünenberger, P.H., Mark, A.E., Querol, E., Avilés, F.X. & van Gunsteren, W.F. (1996). Free energies of transfer of Trp analogs from chloroform to water: Comparison of theory and experiment, and importance of adequate treatment of electrostatic and internal interaction. *J. Am. Chem. Soc.* **118**, 6285–6294

Davis, M.E. & McCammon, J.A. (1990). Electrostatics in biomolecular structure and dynamics. *Chem. Rev.* **90**, 509–521

Dillen, J.L.M. (1995a). An empirical force field. I. Alkanes. *J. Comput. Chem.* **16**, 565–609

Dillen, J.L.M. (1995b). An empirical force field. II. Crystalline alkanes. *J. Comput. Chem.* **16**, 610–615

Dinur, U. & Hagler, A.T. (1991). New approaches to empirical force fields. In: Reviews in computational chemistry, Vol. II, Lipkowitz, K.B. & Boyd, D.B., Eds., VCH Publishers Inc., New York, pp 99–164

Dinur, U. & Hagler, T.A. (1994). On the functional representation of bond energy functions. *J. Comput. Chem.* **9**, 919–924

Elrod, M.J. & Saykally, R.J. (1994). Many-body effects in intermolecular forces. *Chem. Rev.* **94**, 1975–1997

Engelsen, S.B., Fabricius, J. & Rasmussen, K. (1995a). The consistent force field. 1. Methods and strategies for optimization of empirical potential energy functions. *Acta Chem. Scand.* **48**, 548–552

Engelsen, S.B., Fabricius, J. & Rasmussen, K. (1995b). The consistent force field. 2. An optimized set of potential energy functions for the alkanes. *Acta Chem. Scand.* **48**, 553–565

Ermler, W.C. & Hsieh, H.C. (1990). Analytical representation and vibrational-rotational analysis of *ab initio* potential energy and property surfaces, In: Advances in molecular electronic structure theory, Vol. 1, Calculation and characterization of molecular potential energy surfaces, Dunning Jr., T.H., Ed., JAI Press Inc., London, pp 1–44

Essex, J.W. & Jorgensen, W.L. (1995). An empirical boundary potential for water droplet simulations. *J. Comput. Chem.* **16**, 951–972

Fincham, D. (1994). Optimisation of the Ewald sum for large systems. *Mol. Sim.* **13**, 1–9

Fraternali, F. & van Gunsteren, W.F. (1996). An efficient mean solvation force model for use in molecular dynamics simulations of proteins in aqueous solution. *J. Mol. Biol.* **256**, 939

Frenkel, D. (1993). Monte Carlo simulations: A primer. In: Computer simulation of biomolecular systems, theoretical and experimental applications, Vol. II, van Gunsteren, W.F., Weiner, P.K. & Wilkinson, A.J., Eds., ESCOM Science Publishers, B.V., Leiden, The Netherlands, pp 37–66

Field, M.J. (1993). The simulation of chemical reactions. In: Computer simulation of biomolecular systems, theoretical and experimental applications, Vol. II, van Gunsteren, W.F., Weiner, P.K. & Wilkinson, A.J., Eds., ESCOM Science Publishers, B.V., Leiden, The Netherlands, pp 82–123

Figueirido, F., Del Buono, G.S. & Levy, R.M. (1995). On finite-size effects in computer simulations using the Ewald potential. *J. Chem. Phys.* **103**, 6133–6142

Gelin, B.R. (1993). Testing and comparison of empirical force fields: Techniques and problems. In: Computer simulation of biomolecular systems, theoretical and experimental applications, Vol. II, van Gunsteren, W.F., Weiner, P.K. & Wilkinson, A.J., Eds., ESCOM Science Publishers, B.V., Leiden, The Netherlands, pp 127–146

Gerber, P.R. (1992). Peptide mechanics: a force-field for peptides and proteins working with entire residues as smallest units. *Biopolymers* **32**, 1003–1017

Gerber, P.R. & Müller, K. (1995). MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry. *J. Comput. Aided Mol. Design* **9**, 251–268

Hagler, A.T., Lifson, S. & Dauber, P. (1979a). Consistent force field studies of intermolecular forces in hydrogen-bonded crystals. 2. A benchmark for the objective comparison of alternative force fields. *J. Am. Chem. Soc.* **101**, 5122–5130

Hagler, A.T., Lifson, S. & Dauber, P. (1979b). Consistent force field studies of intermolecular forces in hydrogen-bonded crystals. 3. The C=O H-O hydrogen bond and the analysis of the energetics and packing of carboxylic acids. *J. Am. Chem. Soc.* **101**, 5131–5141

Hagler, A.T., Stern, P.S., Sharon, R., Becker, J.M. & Naider, F. (1979c). Computer simulation of the conformational properties of oligopeptides. Comparison of theoretical methods and analysis of experimental results. *J. Am. Chem. Soc.* **101**, 6842-6852

Hagler, A.T. & Ewig, C.S. (1994). On the use of quantum energy surfaces in the derivation of molecular force fields. *Comp. Phys. Commun.* **84**, 131–155

Halgren, T.A. (1990). Maximally diagonal force constants in dependent angle-bending coordinates. 2. Implications for the design of empirical force-fields. *J. Am. Chem. Soc.* **112**, 4710–4723

Halgren, T.A. (1992). Representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules, and vdW parameters. *J. Am. Chem. Soc.* **114**, 7827–7843

Hart, J.R. & Rappé, A.K. (1992a). Van der Waals functional forms for molecular simulations. *J. Chem. Phys.* **97**, 1109-1115

Hart, J.R. & Rappé, A.K. (1992b). Reply to comment on: van der Waals functional forms for molecular simulations. *J. Chem. Phys.* **98**, 2492

Harvey, S.C. (1989). Treatment of electrostatic effects in macromolecular modeling. *Proteins: Struct. Funct. Genet.* **5**, 78–92

Hehre, W.J., Radom, L., Schleyer, P.v.R & Pople, J.A. (1986). *Ab initio* molecular orbital theory, John Wiley & Sons, New York, pp 1-548

Huber, T., Torda, A.E. & van Gunsteren, W.F. (1994). Local elevation: A method for improving the searching properties of molecular dynamics simulation. *J. Comput.-Aided Mol. Design* **8**, 695–708

Hummer, G., Soumpasis, D.M. & Neumann, M. (1992). Pair correlations in an NaCl-SPC water model. Simulations versus extended RISM computations. *Mol. Phys.* **77**, 769–785

Hünenberger, P.H., Mark, A.E. & van Gunsteren, W.F. (1995a). Computational approaches to study protein unfolding: Hen egg white lysozyme as a case study. *Proteins: Struct. Funct. Genet.* **21**, 196–213

Hünenberger, P.H., Mark, A.E. & van Gunsteren, W.F. (1995b). Fluctuation and crosscorrelation analysis of protein motions observed in nanosecond molecular dynamics simulations. *J. Mol. Biol.* **252**, 492–503

Hwang, M.J., Stockfisch, T.P. & Hagler, A.T. (1994). Derivation of class II force-fields. 2. Derivation and characterization of a class II force field, CFF93, for the alkyl functional group and alkane molecules. *J. Am. Chem. Soc.* **116**, 2515–2525

Jones, D.T. (1994). *De novo* protein design using pairwise potentials and genetic algorithm. *Prot. Sci.* **3**, 567–574

Jorgensen, W.L. & Tirado-Rives, J. (1988). The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **110**, 1657–1666

Keith, T.A. & Frisch, M.J. (1994). Inclusion of explicit solvent molecules in a selfconsistent-reaction field model of solvation, In: Modeling the hydrogen bond, Smith, D.A., Ed., American Chemical Society, Washington DC, pp 22–35

Kestin, J., Knierim, K., Mason, E.A., Najafi, B., Ro, S.T. & Waldman, M. (1984). Equilibrium and transport properties of the noble gases and their mixtures at low density. *J. Phys. Chem. Ref. Data* **13**, 229–303

King, G. & Warshel, A. (1989). A surface constrained all-atom solvent model for effective simulations of polar solutions. *J. Chem. Phys.* **91**, 3647–3661

Levitt, M. (1974). Energy refinement of hen egg-white lysozyme. *J. Mol. Biol.* **82**, 393–420

Levitt, M. (1983a). Molecular dynamics of native proteins. I. Computer simulation of trajectories. *J. Mol. Biol.* **168**, 595–620

Levitt, M. (1983b). Molecular dynamics of native protein. II. Analysis and nature of motion. *J. Mol. Biol.* **168**, 621–657

Levitt, M., Hirshberg, M., Sharon, R. & Daggett, V. (1995). Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution. *Comput. Phys. Commun.* **91**, 215–231

Lifson, S. & Warshel, A. (1968). Consistent force field calculations of conformations, vibrational spectra, and enthalpies of cycloalkane and n-alkane molecules. *J. Chem. Phys.* **49**, 5116–5129

Lifson, S., Hagler, A.T. & Dauber, P. (1979). Consistent force field studies of intermolecular forces in hydrogen-bonded crystals. 1. Carboxylic acids, amides, and the C=O···H hydrogen bonds. *J. Am. Chem. Soc.* **101**, 5111–5121

Lifson, S. & Stern, P.S. (1982). Born-Oppenheimer energy surfaces of similar molecules: Interrelations between bond lengths, bond angles, and frequencies of normal vibrations in alkanes. *J. Chem. Phys.* **77**, 4542–4550

Lii, J.-H. & Allinger, N.L. (1989a). Molecular mechanics. The MM3 force field for hydrocarbons. 2. Vibrational frequencies and thermodynamics. *J. Am. Chem. Soc.* **111**, 8566–8575

Lii, J.-H. & Allinger, N.L. (1989b). Molecular mechanics. The MM3 force field for hydrocarbons. 3. The van der Waals potentials and crystal data for aliphatic and aromatic hydrocarbons. *J. Am. Chem. Soc.* **111**, 8576–8582

Lipkowitz, K.B. & Boyd, D.B. (1990ff). Reviews in computational chemistry, Vol. I-VII, VCH Publishers Inc., New York

Liu, H., Müller-Plathe, F. & van Gunsteren, W.F. (1994). A molecular dynamics simulation study with a combined quantum mechanical and molecular mechanical potential energy function: Solvation effects on the conformational equilibrium of dimethoxyethane. *J. Chem. Phys.* **102**, 1722–1730

Liu, H., Müller-Plathe, F. & van Gunsteren, W.F. (1996a). A combined quantum/classical molecular dynamics study of the catalytic mechanism of HIV-protease. *J. Mol. Biol.* **261**, 454

Liu, H., Müller-Plathe, F. & van Gunsteren, W.F. (1996b). Molecular dynamics with a quantum-chemical potential: Solvent effect on an SN2 reaction at nitrogen. *Chem. Eur. J.* **2**, 191

Loncharich, R.J. & Brooks, B.R. (1989). The effects of truncating long-range forces on protein dynamics. *Proteins: Struct. Funct. Genet.* **6**, 32–45

Luty, B.A., Tironi, I.G. & van Gunsteren, W.F. (1995). Lattice-sum methods for calculating electrostatic interactions in molecular simulations. *J. Chem. Phys.* **103**, 3014–3021

Luty, B.A. & van Gunsteren, W.F. (1996). Calculating electrostatic interactions using Particle-Particle-Particle-Mesh method with non-periodic long-range interactions. *J. Phys. Chem.* **100**, 2581

MacKerell Jr., A.D., Wiórkiewicy-Kuczera, J. & Karplus, M. (1995). An all-atom empirical energy function for the simulation of nucleic acids. *J. Am. Chem. Soc.* **117**, 11946–11975

Madura, J.D. & Pettitt, B.M. (1988). Effects of truncating long-range interactions in aqueous ionic solution simulations. *J. Chem. Phys.* **150**, 105–108

Maple, J.R., Dinur, U. & Hagler, A.T. (1988). Derivation of force fields for molecular mechanics and dynamics from *ab initio* energy surfaces. *Proc. Natl. Acad. Sci. USA* **85**, 5350–5354

Maple, J.R., Hwang, M.-J., Stockfisch, T.P., Dinur, U., Waldman, M., Ewig, C.S. & Hagler, A.T. (1994a). Derivation of class II force-fields. I. Methodology and quantum force-field for the alkyl functional group and alkane molecules. *J. Comput. Chem.* **15**, 162–182

Maple, J.R., Hwang, M.-J., Stockfisch & Hagler, A.T. (1994b). Derivation of class II force fields. III. Characterization of a quantum force field for alkanes. *Isr. J. Chem.* **34**, 195–231

Mayo, S.L., Olafson, B.D. & Goddard III, W.A. (1990). DREIDING: A generic force- field for molecular simulations. *J. Phys. Chem.* **94**, 8897–8909

McCammon, J.A. & Harvey, S.C. (1987). Dynamics of proteins and nucleic acids, Cambridge University Press, Cambridge

Momany, F.A. & Rone, R. (1992). Validation of the general purpose QUANTA 3.2/CHARMm force field. *J. Comput. Chem.* **13**, 888–900

Müller-Plathe, F. & van Gunsteren, W.F. (1994). Can simple quantum-chemical continuum models explain the gauche effect in poly(ethylene oxyde)? *Macromolecules* **27**, 6040–6045

Némethy, G., Gibson, K.D., Palmer, K.A., Yoon, C.N., Paterlini, G., Zagari, A., Rumsey, S. & Scheraga, H.A. (1992). Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interaction for use in the ECEPP/3 algorithm, with application to proline-containing peptides. *J. Phys. Chem.* **96**, 6472–6484

Neumann, M. (1983). Dipole moment fluctuation formulas in computer simulations of polar systems. *Mol. Phys.* **50**, 841–858

Neumann, M., Steinhauser, O. & Pawley, G.S. (1984). Consistent calculation of the static and frequency-dependent dielectric constant in computer simulations. *Mol. Phys.* **52**, 97–113

Nilsson, L. & Karplus, M. (1986). Empirical energy functions for energy minimization and dynamics of nucleic acids. *J. Comput. Chem.* **7**, 591–616

Oie, T., Maggiora, G.M., Christoffersen, R.E. & Duchamp, D.J. (1981). Development of a flexible intra- and intermolecular empirical potential function for large molecular systems. *Int. J. Quant. Chem. Quant. Biol. Symp.* **8**, 1–47

Pearlman, D.A., Case, D.A., Caldwell, J.W., Ross, W.S., Cheatham III, T.E., DeBolt, S., Ferguson, D., Seibel, G. & Kollman, P. (1995). AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **91**, 1–41

Pettitt, B.M. & Karplus, M. (1985). Role of electrostatics in the structure, energy, and dynamics of biomolecules: A model study of N-methylalanylacetamide. *J. Am. Chem. Soc.* **107**, 1166–1173

Prevost, M., van Belle, D., Lippens, G. & Wodak, S. (1990). Computer simulations of liquid water: treatment of long-range interactions. *Mol. Phys.* **71**, 587–603

Rappé, A.K., Casewit, C.J., Colwell, K.S., Goddard III, W.A. & Skiff, W.M. (1992). UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **114**, 10024-10035

Russel, S.T. & Warshel, A. (1985). Calculations of electrostatic energies in proteins. The energetics of ionized groups in bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* **185**, 389–404

Schreiber, H. & Steinhauser, O. (1992a). Taming cut-off induced artifacts in molecular dynamics studies of solvated polypeptides. *J. Mol. Biol.* **228**, 909–923

Schreiber, H. & Steinhauser, O. (1992b). Molecular dynamics studies of solvated polypeptides: Why the cut-off scheme does not work. *Chem. Phys.* **168**, 75–89

Schreiber, H. & Steinhauser, O. (1992c). Cutoff size does strongly influence molecular dynamics results on solvated polypeptides. *Biochemistry* **31**, 5856–5860

Smith, J.C. & Karplus, M. (1992). Empirical force field study of geometries and conformational transitions of some organic molecules. *J. Am. Chem. Soc.* **114**, 801-812

Smith, P.E. & van Gunsteren, W.F. (1993). Methods for the evaluation of long range electrostatic forces in computer simulations of molecular systems. In: Computer simulation of biomolecular systems, theoretical and experimental applications, Vol. II, van Gunsteren, W.F., Weiner, P.K. & Wilkinson, A.J., Eds., ESCOM Science Publishers, B.V., Leiden, The Netherlands, pp 182–212

Smith, P.E. & Pettitt, B.M. (1995). Efficient Ewald electrostatic calculations for large systems. *Comput. Phys. Commun.* **91**, 339–344

Smith, P.E. & van Gunsteren, W.F. (1995). Reaction field effects on the simulated properties of liquid water. *Mol. Sim.* **15**, 233–245

Steinbach, P.J. & Brooks, B.R. (1994). New spherical-cutoff methods for long-range forces in macromolecular simulation. *J. Comput. Chem.* **15**, 667–683

Tironi, I.G., Sperb, R., Smith, P.E. & van Gunsteren, W.F. (1995). A generalized reaction field method for molecular dynamics simulations. *J. Chem. Phys.* **102**, 5451–5459

Tomasi, J. & Persico, M. (1994). Molecular interactions in solution: An overview of methods based on continuous distributions of the solvent. *Chem. Rev.* **94**, 2027–2094

Ulrich, P., Scott, W., van Gunsteren, W.F. & Torda. A.E (1997). Protein structure prediction force fields: parametrization with quasi-newtonian dynamics. *Proteins: Struct. Funct. Genet.* **27**, 367–384

van Gunsteren, W.F. & Karplus, M. (1982). Effect of constraints on the dynamics of macromolecules. *Macromolecules* **15**, 1528-1544

van Gunsteren, W.F. & Berendsen, H.J.C (1987). Groningen molecular simulation (GROMOS) library manual, Biomos, Nijenborgh 4, Groningen, The Netherlands

van Gunsteren, W.F., Weiner, P.K. & Wilkinson, A.J. (1989ff). Computer simulation of biomolecular systems, theoretical and experimental applications, Vol. I–II, ESCOM Science Publishers, B.V., Leiden, The Netherlands

van Gunsteren, W.F. & Berendsen, H.J.C (1990). Computer simulation of molecular dynamics: Methodology, applications and perspectives in chemistry. *Angew. Chem. Int. Ed. Engl.* **29**, 992–1023

van Gunsteren, W.F. & Mark, A.E. (1992). On the interpretation of biochemical data by molecular dynamics computer simulation. *Eur. J. Biochem.* **204**, 947–961

van Gunsteren, W.F. (1993). Molecular dynamics and stochastic dynamics simulation: A primer. In: Computer simulation of biomolecular systems, theoretical and experimental applications, Vol. II, van Gunsteren, W.F., Weiner, P.K. & Wilkinson, A.J., Eds., ESCOM Science Publishers, B.V., Leiden, The Netherlands, pp 3–36

van Gunsteren, W.F., Luque, F.J., Timms, D. & Torda, A.E. (1994). Molecular mechanics in biology: From structure to function, taking account of solvation. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 847–863

van Gunsteren, W.F., Huber, T. & Torda, A.E. (1995a). Biomolecular modelling: overview of types of methods to search and sample conformational space. European Conference on Computational Chemistry (E.C.C.C. 1), *American Institute of Physics Conf. Proc.* **330**, 253–268

van Gunsteren, W.F., Hünenberger, P.H., Mark, A.E., Smith, P.E. & Tironi, I.G. (1995b). Computer simulation of protein motion. *Chem. Phys. Commun.* **91**, 305–319

van Schaik, R.C., Berendsen, H.J.C. Torda, A.E. & van Gunsteren, W.F. (1993). A structure refinement method based on molecular dynamics in four spatial dimensions. *J. Mol. Biol.* **234**, 751–762

Vedani, A. (1988). YETI: An interactive molecular mechanics program for small- molecule protein complexes. *J. Comput. Chem.* **9**, 269–280

Waldman, M. & Hagler, A.T. (1993). New combining rules for rare gas van der Waals parameters. *J. Comput. Chem.* **14**, 1077–1084

Wang, L. & Hermans, J. (1995). Reaction field molecular dynamics simulation with Friedman's image charge method. *J. Phys. Chem.* **99**, 12001–12007

Warshel, A. & Lifson, S. (1970). Consistent force field calculations. II. Crystal structures, sublimation energies, molecular and lattice vibrations, molecular conformations, and enthalpies of alkanes. *J. Chem. Phys.* **53**, 582–594

Warshel, A. (1991). Computer modeling of chemical reactions in enzymes and solutions, Wiley-Interscience, John Wiley & Sons, Inc., New York, pp 1-236

Weiner, P.K. & Kollman, P.A. (1981). AMBER: assisted model building with energy refinement. A general program for modeling molecules and their interactions. *J. Comput. Chem.* **2**, 287–303

Weiner, S.J., Kollman, Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta Jr., S. & Weiner, P. (1984). A new force field for molecular mechanical simulations of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**, 765–784

Weiner, S.J., Kollman, P.A., Nguyen, D.T. & Case, D.A. (1986). An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.* **7**, 230–252

Whitnell, R.M. & Wilson, K.R. (1993). Computational molecular dynamics of chemical reactions in solution. In: Reviews in computational chemistry, Vol. IV, Lipkowitz, K.B. & Boyd, D.B., Eds., VCH Publishers Inc., New York, pp 67–148

Wood, R.H. (1995). Continuum electrostatics in a computational universe with finite cutoff radii and periodic boundary conditions: Correction to computed free energies of ionic solvation. *J. Chem. Phys.* **103**, 6177–6187

Yun-Yu, S., Lu, W. & van Gunsteren, W.F. (1988). On the approximation of solvent effects on the conformation and dynamics of cyclosporin A by stochastics dynamics simulation techniques. *Mol. Sim.* **1**, 369-383

Zavitsas, A.A. & Beckwith A.L.J (1989). New potential energy function for bond extensions. *J. Phys. Chem.* **93**, 5419–5426

Zerner, M.C. (1991). Semiempirical molecular orbital methods. In: Reviews in computational chemistry, Vol. II, Lipkowitz, K.B. & Boyd, D.B., Eds., VCH Publishers Inc., New York, pp 313–365